

## BAB V

### PENGUJIAN DAN ANALISIS

Pada bab ini akan dilakukan pengujian dan analisis dari sistem pendukung keputusan keminatan Sekolah Menengah Atas menggunakan metode *modified k-nearest neighbor*. Pengujian dilakukan untuk mengetahui hasil akurasi dari implementasi yang dilakukan.

#### 5.1 PENGUJIAN

Pada pengujian sistem pendukung keputusan keminatan SMA menggunakan 161 *dataset*. Pengujian sistem ini dilakukan dengan nilai  $k=1$  sampai  $k=20$ .

##### 5.1.1 Pengujian Pengaruh Nilai $k$

Pengujian sistem pendukung keputusan keminatan SMA dilakukan untuk mengetahui pengaruh nilai  $k$  terhadap akurasi dimana data yang digunakan yaitu dengan data yang kelas seimbang. Penentuan *dataset* keminatan ditentukan berdasarkan rumus keminatan dan keminatan asli dari pihak sekolah. Dari proses pengujian didapat akurasi rata-rata maksimum yaitu 86.1% pada saat nilai  $k=3$ . Hasil pengujian ditunjukkan pada tabel 5.1 dan tabel 5.2.

Tabel 5.1 Hasil Uji Coba Pengaruh Nilai  $k$  *Dataset* Menggunakan Rumus Keminatan

Nilai $k$	Akurasi Pengujian Untuk Data <i>Training</i> (%)			Rata-Rata Akurasi (%)
	20%	40%	60%	
1	74.6	75.3	74.7	74.87
2	74.6	76.5	75.8	75.63
3	84.6	86.4	87.3	86.1
4	62	64.2	64.8	63.67
5	62	63	63.7	62.9
6	59.2	61.7	62.6	61.167
7	56.3	59.3	60.4	58.67
8	52.1	56.8	58.2	55.7

9	52.1	55.6	54.9	54.2
10	52.1	55.6	57.1	54.93
11	52.1	51.9	58.2	54.07
12	50.7	54.3	56	53.67
13	49.3	53.1	54.9	52.43
14	49.3	53.1	53.8	52.07
15	49.3	50.6	53.7	51.2
16	49.3	53.1	54.9	52.43
17	47.9	51.9	53.8	51.2
18	46.5	50.6	52.7	49.93
19	45.1	48.1	50.5	47.9
20	45.1	49.4	51.6	48.7

Tabel 5.2 Hasil Uji Coba Pengaruh Nilai  $k$  Dataset Asli

Nilai $k$	Akurasi Pengujian Untuk Data <i>Training</i> (%)			Rata-Rata Akurasi (%)
	20%	40%	60%	
1	46.3	45.1	50.9	47.4333
2	43.9	44.6	53.8	47.4333
3	41.5	51.6	50.2	47.7667
4	41.1	48.1	50.5	46.5667
5	39.1	49.4	51.6	46.7

Berdasarkan tabel 5.2 terlihat masih banyak siswa yang keminatannya tidak sesuai dengan kemampuannya. Hal ini disebabkan karena adanya penentuan keminatan siswa yang tidak menggunakan rumus sehingga ada keminatan yang tidak sesuai.

### 5.1.2 Pengujian Pada Jumlah Data Testing Tetap Dengan Jumlah Data *Training* Berbeda

Pada pengujian ini akurasi yang diambil merupakan akurasi terbaik sesuai parameter  $k$  dimana  $k$  terbaik yaitu 3. Pengujian ini dilakukan dengan menambahkan jumlah data *training* dengan kelas seimbang. Hasil pengujian ditunjukkan pada tabel 5.3.

Tabel 5.3 Hasil Pengujian dengan Jumlah Data Testing Tetap dengan Jumlah Data *Training* Berbeda

Jumlah Data Testing	Jumlah Data Training	Akurasi (%)
70	30	70.2
	60	72.8
	90	84.6

### 5.1.3 Pengujian Data *Training* Kelas Seimbang Dan Kelas Tidak Seimbang

Pada pengujian ini membandingkan hasil akurasi pada peningkatan jumlah data *training* kelas seimbang dan kelas tidak seimbang. Hasil pengujian ditunjukkan pada tabel 5.4

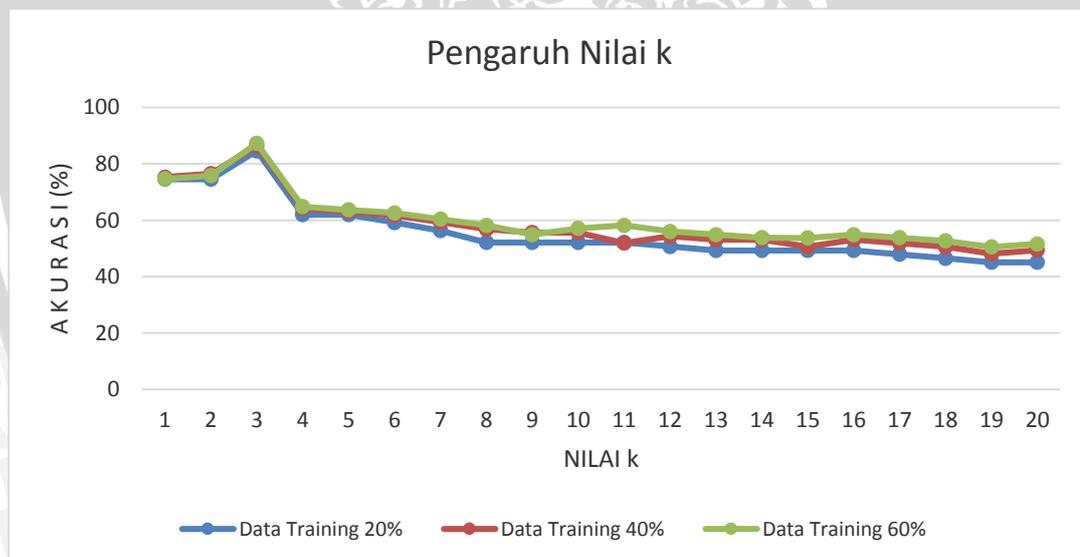
Tabel 5.4 Hasil Pengujian Kelas Seimbang dan Tidak Seimbang

Nilai K	Rata - Rata Akurasi (%)					
	Kelas Seimbang			Kelas Tidak Seimbang		
	Jumlah Data Training			Jumlah Data Training		
	30	60	90	30	60	90
1	74.45	75.15	74.55	59.35	59.95	59.25
2	74.45	76.35	75.65	60.45	61.15	59.25
3	84.45	86.25	87.15	71.95	71.05	69.25
4	61.85	64.05	64.65	49.45	48.85	46.65
5	61.85	62.85	63.55	48.35	47.65	46.65
6	59.05	61.55	62.45	47.25	46.35	43.85
7	56.15	59.15	60.25	45.05	43.95	40.95
8	51.95	56.65	58.05	42.85	41.45	36.75
9	51.95	55.45	54.75	39.55	40.25	36.75
10	51.95	55.45	56.95	41.75	40.25	36.75
11	51.95	51.75	58.05	42.85	36.55	36.75
12	50.55	54.15	55.85	40.65	38.95	35.35
13	49.15	52.95	54.75	39.55	37.75	33.95
14	49.15	52.95	53.65	38.45	37.75	33.95
15	49.15	50.45	53.55	38.35	35.25	33.95
16	49.15	52.95	54.75	39.55	37.75	33.95
17	47.75	51.75	53.65	38.45	36.55	32.55
18	46.35	50.45	52.55	37.35	35.25	31.15
19	44.95	49.25	51.45	36.25	34.05	29.75
20	44.95	47.95	50.35	35.15	32.75	29.75

## 5.2 ANALISIS

### 5.2.1 Analisis Pengaruh Nilai $k$

Berdasarkan hasil uji coba yang dilakukan pada masing – masing data uji, perubahan nilai akurasi rata – rata cenderung semakin menurun dengan adanya penambahan nilai  $k$ . Hal ini disebabkan karena semakin besar nilai  $k$  yang digunakan maka semakin banyak ketetanggaan yang digunakan sehingga kemungkinan untuk terjadi *noise* semakin besar. Semakin kecil nilai  $k$  berarti semakin sedikit jumlah tetangga yang digunakan untuk proses klasifikasi. Metode euclidean distance digunakan untuk mencari jarak terdekat antar data dimana semakin kecil nilainya maka jarak antar data semakin dekat, dengan demikian ketika nilai  $k$  kecil maka hanya tetangga yang memiliki kedekatan data terbaik saja yang digunakan. Pada saat  $k = 3$  dengan nilai akurasi terbaik memiliki kemungkinan nilai ketetanggaan yang terdekat adalah yang terbaik. Grafik pengaruh nilai  $k$  terhadap akurasi dapat dilihat pada gambar 5.1.

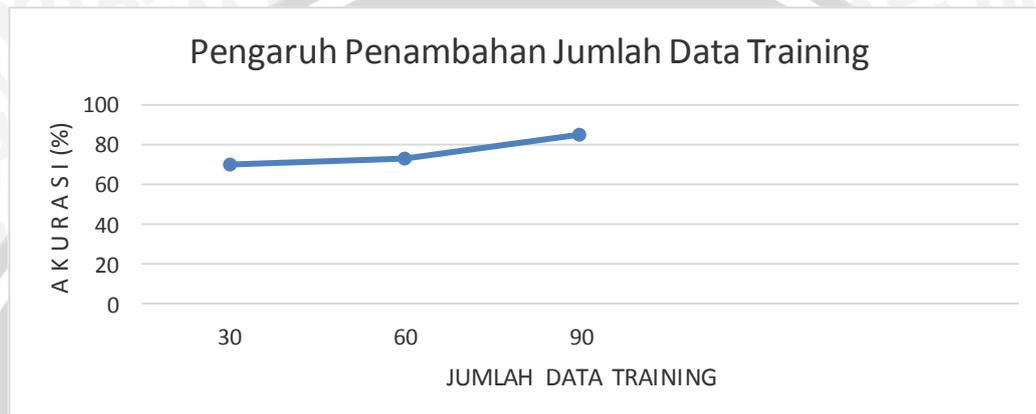


Gambar 5.1 Grafik Pengaruh Nilai  $k$  Terhadap Akurasi

### 5.2.2 Analisis Pengujian pada Jumlah Data *Testing* Tetap dengan Jumlah Data *Training* Berbeda

Berdasarkan uji coba yang dilakukan, terlihat bahwa data *training* berpengaruh terhadap nilai akurasi yang dihasilkan. Peningkatan jumlah data

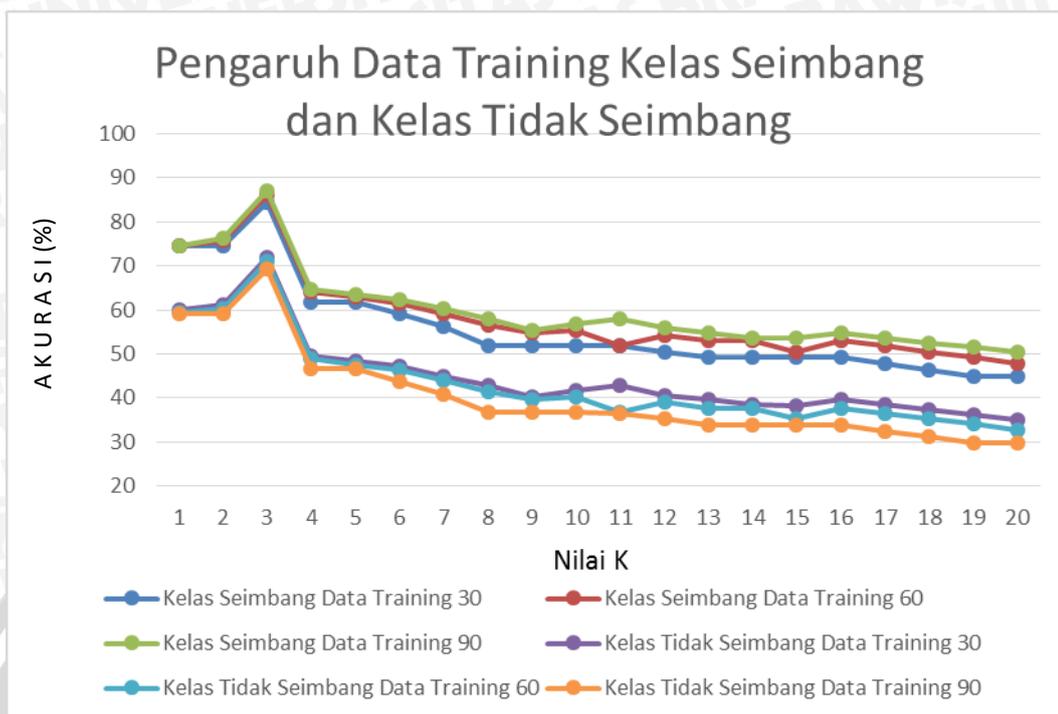
*training* turut disertai dengan peningkatan nilai akurasi. Hal ini disebabkan karena semakin banyak data *training* yang digunakan maka semakin banyak data yang dibandingkan maka semakin banyak jarak yang mendekati. Dalam pengujian ini data *training* yang digunakan dalam keadaan kelas seimbang. Hasil pengujian pada jumlah data *testing* tetap dengan jumlah data *training* berbeda ditunjukkan pada gambar 5.2.



Gambar 5.2 Grafik Penambahan Jumlah Data *Training* Terhadap Akurasi

### 5.2.3 Analisis Pengujian Data *Training* Kelas Seimbang dan Kelas Tidak Seimbang

Berdasarkan hasil pengujian, untuk data *training* dengan kelas tidak seimbang mengalami penurunan hasil akurasi seiring dengan bertambahnya data *training*. Hal sebaliknya terjadi pada data *training* kelas seimbang, peningkatan akurasi terjadi seiring dengan bertambahnya jumlah data *training*. Pada saat data *training* kelas tidak seimbang, terjadi dominasi kelas tertentu yang menyebabkan terjadinya *noise* yang menyebabkan terjadinya kesalahan dalam mengklasifikasi data dan cenderung mengacu pada kelas yang mendominasi. Hasil pengujian data *training* kelas seimbang dan kelas tidak seimbang dapat dilihat pada gambar 5.3.



Gambar 5.3 Grafik Pengaruh Data *Training* Kelas Seimbang dan Kelas Tidak Seimbang Terhadap Akurasi

